DOCUMENT RESUME

ED 150 201 TM 006 950

AUTHOR Brophy, Jere E.
TITLE Training Teachers in Experiments: Considerations Relating to Nonlinearity and Context Effects.
PUB DATE Apr 77
CONTRACT NIE-C-74-0089
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Classroom Research; Correlation; Data Analysis; Educational Research; *Effective Teaching; Elementary Secondary Education; *Experimental Teaching; Observation; *Research Design; Research Methodology; *Research Problems; *Teacher Behavior; Teacher Improvement; Teacher Role; Validity

ABSTRACT

Methods of achieving compromises between the classical laboratory experiment, and the constrictions affecting educational researchers conducting experiments to build upon correlational process-outcome data are suggested. Ecological validity is achieved through partnership relationships between researchers and the teachers implementing treatments, design of treatments to include clusters of related behaviors rather than just one, and provision of complete and detailed information about what to do and why. Simultaneously, a naturalistic analog of experimental control is achieved through judicious selection of research settings, and the capacity for causal inference is retained through collection of data regarding the degree of implementation and short term outcome. Finally, treatments are designed to be as exportable as possible. (Author/MV)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED150201

TM006 950

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jere Brophy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

# Training Teachers in Experiments:

# Considerations Relating to Nonlinearity and Context Effects

Jere E. Brophy

Michigan State University

This paper was presented as part of a symposium on "Rigor and Relevance in Experiments on Teaching in Classrooms" at the annual meeting of the American Educational Research Association, New York, New York, 1977.

Abstract

This paper suggests ways to achieve compromises between the classical laboratory experiment and the constrictions affecting educational researchers who want to conduct experiments to build upon correlational process-outcome data. Ecological validity is achieved through partnership relationships with teachers implementing treatments, design of treatments to include clusters of related behaviors rather than just one, and provision of complete and detailed information about what to do and why. Simultaneously, a naturalistic analog of experimental control is achieved through judicious selection of research settings, and the capacity for causal inference is retained through collection of implementation and short term outcome data. Finally, treatments are designed to be as exportable as possible.

This project was supported in part by the National Institute of Education Contract NIE-C-74-0089, Correlates of Effective Teaching. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by that office should be inferred.

Reviews by Rosenshine and Furst (1973) and by Dunkin and Biddle (1974) concluded that research on teaching effects was beginning to identify teaching process variables that correlated consistently with outcomes. This trend has continued, producing a variety of replicated process-outcome relationships that can provide the bases for experimental studies.

Programatic research that tests hypotheses derived from correlational work, identifies causal relationships, and builds upon these in developing teacher education approaches is needed if teaching is to become the applied science that it can and should be. This paper deals with some of the complexities involved in this enterprise, especially those relating to nonlinearity of, and context effects in, the process-outcome relationships used to generate experimental hypotheses.

Most fundamentally, experimental follow up of correlational process-outcome findings is needed because correlation does not prove causality, and correlational findings alone cannot provide a solid basis for teacher education. Prescriptions based solely on correlations are not complete, intellectually satisfying, or even particularly rational.

One reason is the inherent ambiguity of correlational findings with respect to causality. Another is that even significant correlations usually are moderately strong at best. This means that even if the teaching variables should prove to be causal, more information will be needed before clear and complete prescription will be possible. Should teachers be told to manifest a particular behavior at all times? Probably not. Sometimes it will be irrelevant or unfeasible.

Should they be told to manifest the behavior whenever it <u>is</u> possible? Again, probably not. Positive correlations between a teacher behavior and student outcomes do not necessarily imply that the behavior should be maximized,

and common sense suggests that maximizing is seldom if ever appropriate. Instead, we seek to optimize-manifesting the behavior when it is most appropriate and likely to have positive outcomes, but not otherwise. Being able to do this requires much more information than the knowledge that the behavior correlates with outcomes.

Nonlinearity

These considerations follow from two basic attributes of process-outcome data concerning teacher effects. First, although correlation coefficients remain the most common statistics used to indicate relationships, the real relationships are nonlinear. This has been demonstrated in only a few studies (Brophy and Evertson, 1976; Soar, 1977), but there is every reason to assume that it holds generally, perhaps universally. Second, support for a given relationship varies from study to study. Even for variables that consistently correlate with outcomes, there tends to be a wide range in the strength of relationships in studies with positive results. Also, some studies usually fail to support relationships seen in most of the other studies.

Context Effects

Sometimes, the absence of positive results for a teaching variable that usually correlates significantly with outcomes can be attributed to insufficient sampling, insufficient variation among teachers, unusual measurement or methodology, or other methodological causes. Often, though, negative findings are attributable to context effects on the frequencies or appropriateness of teacher behaviors. Some teacher behaviors are irrelevant in certain teaching contexts, so that observation during these contexts will reveal little or nothing about the effects of these behaviors, because they will not occur. Also, when a teacher behavior occurs in more than one context, it sometimes happens that it is appropriate and

effective in one context but not in the other (Brophy and Evertson, Note 1). These contextual phenomena indicate that generalizations must be limited to the appropriate contexts.

These are the basic reasons that underlie the observed nonlinearity of relationships, as well. Closer analysis of a nonlinear relationship will reveal that the teacher behavior involved is relevant and optimal in certain contexts, relevant but not optimal in other contexts (where some alternative is optimal), and irrelevant in still other contexts. Measuring the behavior without attention to these context differences will yield curvilinear relationships to outcomes, but if one takes context into account, relationships that are more specific and linear will be revealed. These relationships are context-specific and thus not generic to all teaching situations, but they also can be lead to much more specific prescriptions to teachers.

Interpreting Correlations

The prospect of using process-outcome correlations as the basis for experiments on teaching is attractive, partly because it seems to be a way to identify important teaching behaviors and virtually guarantee positive experimental results. However, the inherent ambiguity of correlational data, the fact that most relationships prove to be nonlinear when analyzed carefully, and the fact that most have to be qualified to take context effects into account, make this task much more complex that it seems at first. Furthermore, confining oneself to consideration of linear relationships or relationships that are especially strong will only mask these complexities, and attempts to get rid of them by using box scores to "settle" disagreements will fail. Ultimately, we do not really understand process-outcome relationships until we can explain both why they occur when they they do occur and why they do not occur when they do not.

There are no established paradigms for developing this depth of understanding, but much progress can be made simply by observing instruction, interviewing teachers, and bringing to bear information from other research. Analyses of data that seem to prescribe different behavior in different situations is especially valuable, because such data provide clues to the factors that influence the appropriateness and probable success of particular teacher behaviors. Also, it is instructive to identify the alternatives to a particular behavior of interest when trying to develop hypotheses about why the behavior is or is not effective in particular situations. Consideration of the alternative behaviors available to the teacher in these situations often leads to insights about reasons for the success or failure of <u>each</u> of the alternatives.

Consider the alternatives available to teachers when students do not respond, or respond incorrectly, following a question. First, the teacher must decide whether to provide the correct answer or to continue to question the student in an attempt to get a correct response. If the correct answer is provided, the teacher can do this personally, can call on another student, or can allow other students to call out the answer. If the teacher decides to address another question to the original student, this can be a repetition of the original question, an easier question (either a newer and simpler one or an elaboration of the earlier one that provides clues and thus makes it easier to answer), or a question designed to identify the student's problem rather than get an answer to the original question (asking whether or not the student has heard the question, has read the assignment, or has done preparatory work).

Previous research and logical analyses of these alternatives suggest that some would be optimal in some situations but unwise in others. For example, repeating the question without giving any help might make sense if the student

has not responded, but not if the student has given a wrong answer. Also, giving the answer probably is preferable to calling on other students or letting them call out answers, but not necessarily preferable to asking the student another question in an attempt to elicit a response. Giving the answer probably is most appropriate when the teacher is introducing new material or when the student seems completely confused, while attempting to elicit a response through repeated questioning seems more appropriate when there is reason to believe that the student knows the answer or could figure it out with help.

Developing Hypotheses

Distinctions like these elaborate knowledge about the situational differences that influence (or should influence) teacher behavior, and they place teacher behaviors in contexts that specify antecedent conditions and short run outcomes related to goals. The result is the development of a set of related hypotheses concerning when particular teacher behaviors are optimal and why they succeed when they do. Taken together, such hypotheses provide systematic explanations for process-outcome relationships, and they allow specific and differential predictions about what will happen if particular alternatives are used in particular situations.

This makes it possible to train teachers specifically and exhaustively, and to test sophisticated differential hypotheses relating alternative behavior to short term and long term outcomes.

Treatments as Clusters of Hypotheses

An important implication here is that teaching behaviors manipulated in experiments on teaching usually should be considered in related clusters rather than in isolation. The classical experiment involves systematic manipulation of

a single independent variable combined with precise control over all other variables known to be relevant. Even if this were feasible in education, and it seldom is, it probably would not be appropriate. We have to sacrifice this kind of precision and control in order to achieve ecological validity (which exists when the experimental situation is near enough to naturalistic teaching to allow us to generalize experimental results to the classroom). Also, given the generally primitive development of outcome measures and the difficulty of measuring change, it often is necessary to manipulate a cluster of related teaching variables simultaneously in order to create treatments that are powerful enough to have detectable effects.

This need not be a flaw in the experimental design or an impediment to causal inference if experiments are designed so that implementation data and information about short term outcomes are collected. Implementation data concern the degree to which the experimental treatment is being implemented by the teachers in experimental groups, and also the degree to which the treatment is being "implemented" naturalistically by the teachers in control groups. This information is important in educational research, because complete control over teaching behavior usually is neither possible nor desirable.

Complete control usually is not possible because treatment implementation depends upon the experimental teachers, who may forget or be unable to behave as directed at all times or may deliberately substitute for the treatment at times when it seems inappropriate. Even when full control of the treatments used in each group is possible, it usually is not desirable because it creates artificial situations in all groups. The practical questions of interest in educational research concern comparisons of proposed improvements with common practice, not comparisons of hypothesized ideal treatments with their opposites. Treatments based on clusters of related behaviors rather than isolated single

behaviors help facilitate this in two ways. First, they increase the chance that effects will be practically significant as well as statistically significant. Second, they are more acceptable intellectually and emotionally to teachers, and for that reason are easier to implement consistently, than treatments confined to a single behavior.

Implementation and short term outcome data allow one to test predictions concerning each separate behavior involved in the treatment. If the contexts in which each behavior is relevant are clearly spelled out, and if the short term outcomes of performance of the desired behavior are predicted (improved attention, improved response, ability to complete a task with out seeking help, etc.), teachers can be coded for implementation of the treatment in the relevant contexts, and the short term outcomes following successful or unsuccessful implementation can be coded as well. If there is sufficient variation in teacher behavior (uneven implementation in the treatment group and/or some naturalistic implementation in the control group), it will be possible to determine whether or not the behavior in question was systematically associated with long term outcome measures. Again, this can be done for both the treatment and the control groups, assuming sufficient variance.

These design features strike a compromise between the controlled laboratory experiment and the need for ecological validity to provide a basis for generalizing the findings. The interrelationships among the treatment elements and the rationales that underlie and tie them together make the treatment both more powerful and more face valid for teachers, yet the design allows for specific evaluation of each treatment element in addition to assessment of the effectiveness of the treatment as a whole.

Choice of Research Settings

It is important to find a research setting that will provide an appropriate, and if possible optimal, opportunity to test the hypotheses involved. Sometimes the hypotheses point to a particular setting, as when they deal with small group instruction or instruction in particular subject matter at a particular grade level. Even where no such specification exists, however, and even where the treatment is believed to generalize to all "teaching," some settings will be better than others. This is because certain assumptions are built into treatment hypotheses even if the investigators are not aware of them.

One of these concerns the classroom atmosphere. Unless the treatment deals with classroom management, hypotheses relating teacher behavior to outcomes probably assume motivated or at least cooperative students who pay attention to instruction and apply themselves in working on assignments. Where this is not the case, the assumed treatment will not operate no matter what the teacher does.

An appropriate match between the objectives and materials of the curriculum and the present abilities and interests of the students usually is assumed, as well. If the tasks involved in the treatment are much too easy, ceiling effects will minimize variance in outcomes and thus make it difficult or impossible to test the hypotheses. Alternatively, if the tasks are much too difficult, even well implemented treatments that might have been successful under other conditions will not yield enough learning to make a detectable difference. This danger is quite real whenever "grade level" tasks are used in schools or classrooms where most of the students are functioning at levels significantly below grade level.

There usually are implicit assumptions about teachers, as well. Typically, successful implementation means not only that the teachers perform the experimental behaviors as instructed, but also that they be generally competent in handling the

broad range of tasks that teachers must be able to do. At minimum, this would seem to require experienced teachers, and perhaps teachers experienced with the particular grade levels and curricula involved. Even among these teachers, it may be necessary to exclude those who are inept as classroom managers or otherwise deficient in ways that cause their classrooms to differ fundamentally from most others and preclude meaningful tests of hypotheses about instructional techniques.

An implication of these considerations about research context is that experiments designed to test causal hypotheses based on process-outcome correlations should be conducted in ideal settings, at least initially. The teachers should be experienced, competent, and committed to implementation of the treatment, and the students should be appropriate in demographic characteristics, matched to the content and difficulty level of the curriculum, and socialized to the student role. This will have the effect of maximizing the chances for positive results, which is as it should be when the testing of hypotheses about the treatment is the primary interest. In effect, selecting the ideal setting in which to conduct such research is a way to control for other factors that might interfere with treatment implementation or reduce the variance in learning outcomes to a point where hypothesis testing would be impossible.

Chaotic settings should be avoided in early experimental work designed to establish treatment effectiveness. Later, after the effectiveness of the treatment has been established in ideal settings, attention can be focused on the classrooms of student teachers, teachers who cannot control the class, teachers who must deal with drastic variations in student ability and achievement levels, teachers facing entire classrooms full of alienated low achievers, and other classrooms where there may be less chance for the treatment to work.

Training the Teachers

In experimental tests of teaching principles, the teachers should be cooperating partners with the experimenter, not "subjects" given minimal information. Researchers concerned about Hawthorne effects and experimenter bias effects sometimes give teachers minimal information about what to do and/or why to do it. This approach probably does more harm than good. It reduces cooperation and good will, confuses the teachers, minimizes implementation, and introduces artificiality. This works against the goal of getting teachers to implement the treatment both naturalistically and well, in order to provide a credible basis for generalization.

A better way to avoid Hawthorne and experimenter bias effects is to have two or more experimental groups implement different treatments. If specific differential hypotheses are developed and short term outcome data are collected, it will be possible to see if each group implements its treatment successfully and gets systematically good results on outcomes related to its treatment. This design not only equalizes Hawthorne effects and experimenter bias effects, but it provides good assurance that the treatment effects of interest to one group are not of interest to the other.

Whether or not there is more than one experimental group, it is possible to have control groups that do not know about the treatment groups and can be monitored for naturalistic implementation of treatments. Teachers in these groups can be motivated to cooperate in much the same way that they would be motivated if involved in a purely naturalistic process-outcome study. They would be participating in research and getting just as much attention from the experimenter, but they would not be asked to implement specific treatments and thus would not have their attention called to the behaviors of interest. The possibility of experimenter bias effects could be assessed by analyzing the short term outcomes in these control classrooms and those of the experimental classrooms.

This would involve comparing the intensity, enthusiasm, and other qualitative aspects of behavior observed when situations relevant to the treatment principles arose, and, more importantly, comparing the short term outcomes of the behavior of these two teacher groups when they did and did not implement the treatment.

Treatments themselves should be face valid and well articulated. In training the teachers, it is important to circumscribe the contexts for decision and action, identify relevant and approved strategies for each context, specify alternative strategies if more than one way to operate is available, and explain about when and why each alternative is best or not advised. Context information that tells when a strategy is relevant may be just as important as explanations about what the strategy involves and how it should be implemented. Teachers' abilities to recognize contexts that are appropriate and remember the relevant strategies will facilitate their ability to implement the treatment properly.

If sequences of complex behavior are involved, the treatment should include the whole sequence, complete with alternative behaviors relevant at different stages. Instructions should cover exceptions to the general rules, along with the rules themselves. Again, it is essentially irrational to present teachers with partial information and ask them to implement certain behaviors at all times when it really is recommended for certain (unspecified) times ("Be indirect," "Ask complex questions"). Teachers should know the general principles behind prescriptions, not just the prescriptions themselves, so that they can adjust appropriately to situations that the prescriptions do not cover.

Short Term Outcomes

The importance of short term outcomes must be stressed. To monitor treatment implementation, it will be necessary to develop ways to convert behavioral objectives into percentage scores reflecting the implementation of decision rules (as opposed

to simple frequency or rate measures that give information about how often a behavior was observed but do not say whether its appearances were appropriate or not).

Information about the linkages between treatments and short term outcomes allows inferences about how treatments work in addition to documenting that they do work. This is useful whether or not information about long term outcomes is available. In fact, it is more prescriptive and scientifically convincing than data relating to long term outcomes. We probably have placed too much emphasis on long term outcomes, especially the kinds of omnibus achievement tests that provide a credible global indication of how students are progressing but do not allow specific linkages between their progress and anything that the teacher does. Such outcomes may have been overemphasized to the point that worthwhile information about teaching has been ignored because it has not been supported with long term outcome data.

We should retain our concern with outcomes, especially achievement outcomes, as the criteria for effective teaching, but we need to improve upon our use of omnibus achievement tests. We need more delineation of specific differential outcomes hypothesized to result from specific differential teaching behavior, along with tests of hypotheses through well designed experiments that yield credible data. Sometimes, differential predictions will have more to say about the quality or type of achievement than the amount, but this is as it should be. Before decision makers can rationally evaluate alternative teacher behaviors, it will be necessary to document systematic relationships between these teacher behaviors and their contrasting outcomes. Most of this will involve the use of short term outcome data.

Treatment Exportability

Although it is important to make the treatment as complete and credible as possible in the sense described above, it also is important to hold the duration, expense, and general intensity of the treatment to a minimum. This is recommended not so much for scientific reasons as for practical reasons relating to treatment exportability. The more thorough the treatment, the more likely systematic implementation, but the less generalizable to ordinary teaching.

In view of this, it probably is best to start, where feasible, with minimal treatments that involve little external help in the form of supervision, videotaped feedback, or other time consuming or expensive features. If such treatments work (implementation is good), and if they produce desirable effects, they can be exported quickly and cheaply. If not, more high powered treatments can be tried out in subsequent work (although the problem may be lack of specificity, clarity, or even correctness of the treatment itself rather than difficulty of getting teachers to master it or to implement it).

## Summary

This paper has suggested ways to achieve compromises between the classical laboratory experiment and the constrictions affecting educational researchers who want to conduct experiments to build upon correlational process-outcome data. Ecological validity is achieved through partnership relationships with teachers implementing treatments, design of treatments to include clusters of related behaviors rather than just one, and provision of complete and detailed information about what to do and why. Simultaneously, a naturalistic analogue of experimental control is achieved through judicious selection of research settings, and the capacity for causal inference is retained through collection of implementation and short term outcome data. Finally, treatments are designed to be as exportable as possible.

Reference Note

Brophy, J. & Evertson, C. Context effects on classroom process variables. Report No. 76-10, Research and Development Center for Teacher Education, University of Texas, 1976.

References

Brophy, J., & Evertson, C. Learning from teaching: A developmental perspective. Boston: Allyn & Bacon, 1976.

Dunkin, M., & Biddle, B. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In Travers, R. (Ed.). Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Soar, R. An integration of findings from four studies of teacher effectiveness. In Borich, G. (Ed.). The appraisal of teaching: Concepts and process. Reading, Mass.: Addison-Wesley, 1977.